DOCUMENT RESUME

ED 304 455 TM 012 817

AUTHOR Postlethwaite, T. Neville
TITLE The Use of Standardized Tests in Secondary Schools in Four European Countries.
INSTITUTION National Center on Effective Secondary Schools, Madison, WI.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Nov 86
GRANT OERI-G-86-0007
NOTE 38p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Achievement Tests; *Educational Policy; Foreign Countries; *Intelligence Tests; National Competency Tests; National Programs; Secondary Education; *Secondary Schools; Secondary School Teachers; *Standardized Tests; Testing Programs; *Test Use
IDENTIFIERS England; *Europe; Netherlands; Sweden; West Germany

ABSTRACT

The use of standardized testing in secondary schools (with students between 10 and 19 years old) is described for four European countries: (1) England; (2) West Germany; (3) the Netherlands; and (4) Sweden. In the decentralized English system, several published standardized tests are available; they are used less at the secondary level than in primary grades. Tests are used for special needs assessment and educational guidance and there is some trend toward increased use of graded objectives tests. In West Germany, education is the responsibility of the states, rather than the national government. Standardized tests are not used on a population-wide basis, and the use of standardized tests is largel restricted to counseling centers and similar specialists in the schools. Neither achievement nor intelligence tests are often used in the schools. The Netherlands created a national curriculum development center in 1975 and has created national examinations, although they are not yet widely used. Achievement tests are used by teachers only, and intelligence test use is similar to that in West Germany. In Sweden, national standardized tests based on objective techniques are used above the primary levels. In summary, it was generally found that teachers do not use standardized tests of their own accord, mostly because tests are not tailor-made for what the teachers have been teaching. Appendices present three papers summarizing recent developments in England and two aspects of the Swedish testing process. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

F. M. NEWMAN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED304455

The Use of Standardized Tests in Secondary Schools in Four European Countries

T. Neville Postlethwaite
Institute of Comparative Education
Sedanstr. 19
2000 Hamburg 13
Federal Republic of Germany

November 1986

This paper was prepared for the National Center on Effective Schools, University of Wisconsin-Madison, School of Education, which is supported in part by a grant from the Office of Educational Research and Improvement (OERI-G-86-0007). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of this agency or the US Department of Education.

TM012817

The Use of Standardized Tests in Secondary Schools in Four European Countries

T. Neville Postlethwaite

In this paper, standardized tests comprise any tests that are generated outside the school and that are administered in a common fashion to students in a variety of schools. Secondary school means, roughly speaking, schools having students beginning with age 10 - 13 up to an ending age of 16 - 19; since the use of tests varies somewhat according to type of secondary school, a description will be given of the school systems in the four systems for which data are presented. These are England, Federal Republic of Germany, the Netherlands, and Sweden.

In the United States standardized tests are sometimes administered but the results are rarely used by school personnel since the content of the tests rarely bears any relationship to the curriculum, in a particular subject-matter, of the classroom or school. But, in the United States, the system of education is very decentralized. What happens in other countries, both centralized and decentralized? Are tests valid for what is taught in school? How do school personnel use tests or don't they? Do tests "form" what is actually taught in school?

England is a decentralized system of education. It was only in 1986 that it was officially mooted that there should be any form of national curriculum. National examinations are organized in England by several different examining boards and they do influence what is taught in the schools from age 15 onwards; but these are not standardized tests.

In the Federal Republic of Germany, education is the responsibility of each state and not of the national authorities. Each state produces a 'Lehrplan' (or syllabus) for each subject matter for each grade level. There are eleven states in Germany. Only four have been selected: Hamburg (HH), Lower Saxony (LS), Northrhine Westfalia (NRW), and Schleswig-Holstein (SH).

The Netherlands is a relatively decentralized system with regional boards of education deciding on which 'models' of

curriculum should be adopted and each school is then responsible for the implementation. Norm-referenced tests are used for promotion from grade to grade (Nijhof and Streumer, 1985).

Sweden has a national 'Läroplan' (syllabus) for each subject area and text books are produced by publishing companies, but each teacher may add to the curriculum as he / she wishes. There are no formal examinations, although, in the Gymnasium there is continuous assessment.

This article will deal with each of the four countries in turn. Within each country section the school system and secondary education, in particilar, will be described. Secondly, a description of what sort of standardized tests exist will be attempted. Thirdly, an assessment will be made of who uses the tests for which purposes. A general conclusion will be made at the end of the four country descriptions.

A. ENGLAND+

1. The System.

Compulsory schooling lasts from age 5 to age 16. Secondary schooling normally refers to the period from age 11 to age 18. Education is administered by over 100 local education authorities (LEAs) sometimes referred to as 'authorities'.

A widely used convention to identify secondary school year-groups is First years (11.00 - 12.00), Second years (12.00 - 13.00), Third years (13.00 - 14.00), Fourth years (14.00 - 15.00) and Fifth years (15.00 - 16.00). Subsequent year groups are referred to as 1st year Sixth, 2nd year Sixth and 3rd year Sixth, although, few students stay for a 3rd year.

Examinations taken at the age of 16.00 plus years have an important place (a matter of some controversy) in English secondary schools. Historically, these were provided for a minority of pupils in 'grammar schools', who were selected at age 10 by procedures using either test results, head-teacher's recommendations or combinations of these. These examinations for 16 year olds provided a screen for university entrance in that successful

---

+ The information in the section on England was provided by Dr. Ray Summer of the National Foundation for Educational in England & Wales.

candidates were 'matriculated'. Nowadays, the purpose is supposedly to recognise attainments at the end of statutory schooling; but usually, these examinations are regarded as producing qualifications of a general nature. About 90 percent of the fifth year cohort attempt one or more of the examinations, but there are several examining bodies and a multiplicity of examination titles. Consequently, though there is a common structure of grades or levels, there is no formal equating to estimate the equivalence of standards between examinations with the same title provided by different bodies, or examinations set for different subjects by the same body. Though these examinations are seen by the public and many educationists as standard setters, they are not particularly akin to standardised tests because (i) fresh questions are written each year, (ii) there is little pre-testing or prior item analysis, (iii) whilst the multiple-choice parts of the examination may be analysed post hoc, emphasis lies on total (parts added to give whole) scores and their distributions in relation to grade boundaries, and (iv) there is little or no normalisation or reference to a model. Booth (1985 p. 5356) when describing the system of education in the United Kingdom, stated "In the United Kingdom, there is no nationally determined curriculum. However, the examination boards which control the G.C.E. (or its equivalent) exert something of a unifying influence on secondary schools in their area."

The currently diverse sets of examining bodies (8 university-based boards for G.C.E. and 14 regional boards for C.S.E.) have now been formed into examination consortia to deliver a new system of school examination for 16+ pupils, called the General Certificate of Secondary Education. This offers an eight grade structure with the lower levels accessible to pupils whose abilities or attainments are relatively limited. The G.C.S.E. courses started in September 1986 with the first examinations taking place in 1988.

2. <u>What standardized tests exist?</u>

An achievement test in England is called an attainment test and an intelligence test is called an ability test.

No data exist on the tests used in schools. There are two main publishers of tests: NFER - Nelson and Hodder and Stoughton. In 1983 Gipps <u>et al</u> reported a small study in 40 secondary schools

that they undertook on the use of tests. Tables 1 and 2 report the results.

Table 1 :Tests used with different years in 40 secondary schools (frequency of mention in brackets)

Reading	*Mathematics*	*English & Spelling*	*Other*
First Year			
Burt (1)	Bristol Achievement (1)	*Spelling*	NFER VR EF (1)
Daniels & Diack (4)	NFER DE (1)	Daniels & Diack (1)	NFER NV 3 (1)
GAP (2)	NFER ? (1)	Schonell (2)	Nelson Basic Skills (1)
Holborn (4)	Nelson Profiles (1)	SPAR (1)	Nelson CAT (1)
NFER DE (1)	Richmond (4)		Richmond Work Study (1)
NFER ? (1)	Tennel (1)	*English*	Richmond Basic Skills (1)
Neale (1)		NFER EF (1)	Swedish Language Test (1)
Richmond (2)		NFER Language (1)	*Essential IQ (1)
Schonell GWRT (2)		Richmond (3)	
Schonell Silent B (1)			
Vernon B (1)			
Widespan (2)			
Second Year			
Gapadol (1)	Nelson Profiles (1)	Richmond English (1)	Richmond Basic Skills (1)
Neale (1)	Richmond (3)		Richmond Work Study (1)
Richmond (2)			
Third Year			
Daniels & Diack (1)	Richmond (4)	Richmond English (2)	Richmond Basic Skills (1)
Gapadol (1)		Vernon Spelling (1)	Richmond Work Study (1)
Neale (1)			
Richmond (2)			
Widespan (1)			
Fourth Year			
Gapadol (1)			

NB. In the fourth year one LEA reading test was used. In the fifth year one LEA test of numeracy.
* There is some uncertainty about the exact identity of this test.

Table 2 :Tests used diagnostically in 40 secondary schools (frequency of mention in brackets)

Reading	*Spelling*
Burt (1)	Blackwell Spelling Workshop (1)
Carver (1)	Daniels & Diack (1)
Daniels & Diack (10)	Dorcan (1)
Domain Phonic (2)	Margaret Peters (1)
Edinburgh (1)	Schonell (4)
GAP (2)	Swansea (1)
Gapadol (3)	Vernon (1)
Gibson's Phonic (2)	
Holborn (9)	
Jackson Getting Reading Right (1)	*English*
Jackson Phonic (2)	
NFER ? (1)	Daniels & Diack Comprehension (3)
Neale (9)	NFER English Progress (1)
Schonell GWRT (11)	NFER English Comprehension (1)
Schonell Silent Reading (3)	Schonell Diagnostic English (1)
SPAR (3)	
Widespan (4)	
Young (1)	
Mathematics	*Other*
Computational Skills Development Test (1)	Aston Index (7)
NFER ? (1)	Bristol Achievement (1)
Schonell Four Rules (1)	Bristol Social Adjustment (1)
Sproad (1)	NFER VR (1)
Unspecified (3)	Nelson CAT (3)
	Oxford Modern Language (1)
	Raven's Matrices (1)
	Schonell IQ (1)
	Young's NRIT (1)
	WISC (1)
	Unspecified Non-verbal IQ (1)

These tables, from only small undefined samples, illustrate that the primary forms for testing are English and Mathematics, principally in the 1st secondary year. Similarly, the limited amount of abilities testing is done mainly with 1st year pupils. Many of the tests, especially of reading, are extremely dated, i.e. Burt, Schonell, Holborn, but give reading ages, which teachers believe they understand and find useful. The use is principally to identify or confirm poor reading ability and to aid in the placement of children in classes where special provision is made. This use also applies to the Maths tests and the ability tests. The latter, however, are probably more widely used to allocate pupils to bands (i.e. three supposedly hierarchical groups of low, middling and higher ability, not necessarily equal in size), or to allocate children to 'mixed ability classes' of roughly equivalent ability distributions.

The comments made above indicate that tests and uses are linked when the aspect of validity is considered. The majority of the tests named in the tables are normed and standardised; i.e. are norm-referenced. A few of the reading tests purport to be diagnostic (e.g. Neale, passages are read aloud by the pupil and the teacher makes an error analysis), but this claim is difficult to substantiate.

Since the book survey was carried out there have been a number of new tests published. These are noted below, together with some of the others cited which are currently in widespread use. It is not possible to comment on content validity in the majority of cases, as the manuals do not generally address this matter.

Some commonly used tests are:

NFER English Progress: a series for different year groups up to the age of 14-15 years; content is considered dated by English teachers and norms are about 15 years old. Used to check on attainment on entry to secondary school or 'progress' thereafter.

NFER Mathematics Attainment: similar to above.

NFER Basic Mathematics: tests in this series were normed up to 10 years ago; each has item content grid and scoring is taken to indicate areas of competence demonstrated by individual children; total score is normed.

Richmond Test of Basic Skills: an anglicised version of the IOWA Test of Basic Skills; normed after trials in Richmond in 1972 and recently re-normed. In middle secondary years, used as a guide to probable external examination success in certain courses, with quite a lot of credence given to the Study Skills sub-scales.

Bristol Achievement Tests: used as progress checks in basic curriculum areas; similar to Richmond in content and use.

Edinburgh Reading Tests: four stages, with two suitable for secondary schools; sub-tests include Skimming, Vocabulary, Reading for Facts, Point of View and Comprehension.

Cognitive Abilities Test: the well-known American test, somewhat anglicised and with recently completed second standardisation for age 8 through 15. Used to appraise ability (and the dubious concept of potential), for slower learner identification; for banding or mixed ability grouping. Recently, this test has been used by an authority to decide how many teachers will be allocated over and above a standard level, to cope with 'special needs'.

London Reading Test: for use at point of transfer to secondary school or on arrival; normed for Autumn population (October); as progress check and indicator of pupils experiencing reading difficulty.

Profile of Mathematics Skills: level 2, 10-15 year olds; Addition, Subtraction, Multiplication, Division, Operations, Measurement, Money, Fractions, Decimal fractions, Percentages and Diagrams. Though the tests are said to be criterion-referenced and are normed, the main use is supposed to be diagnostic (strengths and weaknesses) as indicated by the Profile. However, subscale reliabilities are moderate (around 0.8) and so differences are likely to be over-interpreted.

Children's Abilities Scales: for 11-12 year olds; sub-tests for Verbal, Non-verbal (symbolic reasoning) and Spatial; standardised in 1983; used for appraisal of secondary pupils' abilities on transfer from primary; placement in groups; educational guidance.

Educational Abilities Scales: five parts; Spatial reasoning, Clerical, Symbolic reasoning, Science reasoning, Mechanical comprehension; for 3rd year pupils, use for educational guidance re. choosing optional courses for external examinations; normed in 1983; unusual answer-until-correct presen-

tation (pupils remove a latex film from multiple-choice response alternatives until correct one appears) which is virtually self-scoring.

Chelsea Diagnostic Mathematics Tests: not normed or referenced to criteria of performance kind; these tests enable pupils to be classified into levels of understanding and characteristic error groups; age-range 12 to 15+; for Algebra, Fractions, Graphs, Measurement, Number operations, Place value and Decimals, Ratio and Proportion, Reflection and Rotation, Vectors.

Reliabilities generally run from KR20 values of 0.96 (for single-age reasoning or attainment tests) to around 0.8 for test sub-scales. Usually, there is little validation data though in recent tests in some manuals studies may be cited and factor analysis (or similar) results quoted.

Some recent developments about graded tests are presented in Appendix I.

3. Who uses the tests for which Purposes?

The published standardised tests are used at the secondary stage far less than at the primary stage. Some uses have been mentioned above but two further aspects are worthy of notice. These are:

(a) Special needs assessment: Legislation in 1981 obliges every authority to implement a policy of providing for pupils' special needs, when assessed as handicapped in some way or as 'below average' (this is generally interpreted as in the lowest 20 percent by general attainment, following the Warnock Report estimate). Teachers were seen as the first in the assessment line, so test results 'on file' are a defensible way of making an appraisal of groups of pupils and proceeding to further multi-professional assessment.

If a pupil, who has been 'satisfactory', suddenly begins to perform poorly then the scores 'on file' are referred to. Occasionally, scores are also used for determining to which learning group a pupil is allocated. In some cases, local education authorities use tests (math or intelligence) to determine how many of the pupils in a school are in the lower half of the achievement or ability range in order to allocate extra teachers to the school.

(b) Educational guidance: A critical point in a pupil's career is choosing between subject options at the age of 14+. School sometimes assesses abilities with standardised tests as well as performance in school subjects. Test scores are then used to counsel pupils and their parents. In recent years, an increasing number of authorities have provided the tests and paid for scoring services.

Further to these two purposes, local education authorities want to know how the pupils in their authority compare in achievement with the nation as a whole. The L.E.A. advisors sometimes want specific school scores for schools in need of extra resources, and sometimes individual pupil scores for allocation to group (or special schools) purposes. In the Inner London Education Authority various indicators, including intelligence test scores, are used as predictors to identify schools well above or well below the regression line (in particular the latter resulting in a visit from an inspector).

There are also national surveys (in Math, English, Science, Modern Languages, and Craft, Design and Technology) conducted by Assessment Performance Units. These are similar to the NAEP surveys in the United States.

As a closing remark, it should be stated that

i) it looks as though the graded objectives tests look as though they will spread more widely since they offer a close match with the curriculum and often involve a pupil's teacher, and

ii) there is a trend towards accrediting some teachers as competent assessors or as supervisors of assessment schemes.

But, despite the above there is still relatively little use of standardized tests in English secondary schools.

B. FEDERAL REPUBLIC OF GERMANY +

1. The System.

Education in the F.R.G. is a state and not a national responsibility. There are eleven states. Each state sets its own syllebuses (Lehrpläne). Publishing companies produce textbooks. These textbooks are then adopted (put on official lists) or not by each state on the grounds of "fitting to" the official syllabus and being well written.

+ Information for this section was provided by Professor Dr. Dr. Rainer Lehmann of Hamburg University. Those who supplied information to Professor Lehmann are listed on page 12.

The state educational authorities organize school inspections that supervise the school-administered examinations and the teaching. There are no nation-wide standardized examinations.

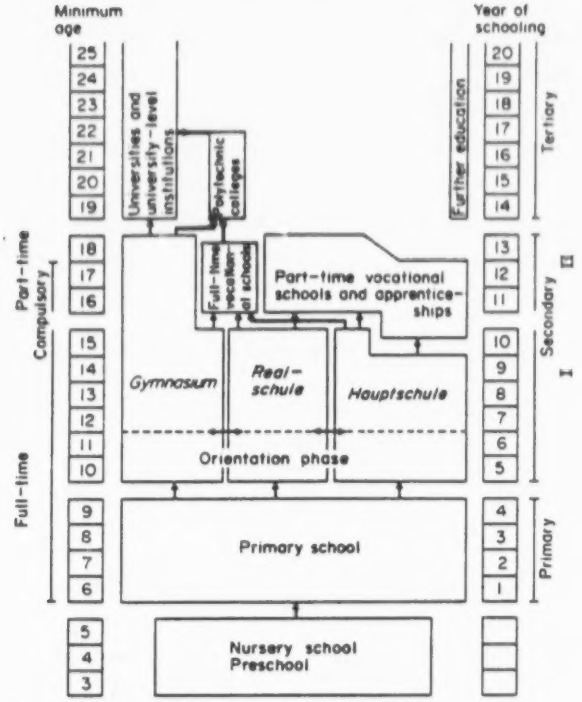


Figure 1
Formal educational system (1980)

As can be seen from Figure 1 secondary school runs from age 10 to 18. There are several school types: Gymnasium, Realschule, Hauptschule and various vocational schools. There are also a handful of so-called comprehensive schools; so-called because they often do not comprize all children in an area but have had the "better" children creamed off into a Gymnasium. There is a different syllabus for each school type in each state. For vocational schools, the state Chamber of Commerce participates in the specification and supervision of examinations.

2. What tests exist?

About 10 publishing companies, the most important being Beltz, Hogrefe and Klett-Cotta, produce achievement tests in core subject areas mostly for the age-group 10 to 15. In some cases, the reliabilities have values above 0.9. Ideally, the test content matches the subject matter covered by the existing

textbooks. However, it is usual for teachers not only to use a textbook but also to produce a good deal of written text material (xeroxed) themselves so that the publishing house tests are not valid and teachers rely on classwork, homework and teacher quizzes for formative and summative evaluation purposes.

A number of verbal and non-verbal intelligence tests exist but only specially trained teachers will use them.

Table 3 on page 10a presents the number of tests that existed in 1984 for Grades 7 and above.

3. Who uses the tests for which purposes?

a) For overall assessment. Standardized tests are not used on a population-wide basis, and they are never employed for assessment purposes at the ministry level. In a few cases, the ministry prescribes a particular test if the school decides to test. In these cases, the ministry has also defined which groups are authorized to administer the tests and to interpret the results. In general, the test-administrators are psychologists from psychological counselling centers or specially trained teachers ("Beratungslehrer"; in Hamburg also "LRS-Lehrer" - LRS = Lese-Rechtschreib-Schwäche = Dyslexia, "Testlehrer") who have taken in-service training or special courses and whose activities are usually taking place in close coordination with counselling centers. In general, the use of standardized tests is largely restricted to counselling centers and similar specialists in the schools.

b) Intelligence tests. In Hamburg, schools are allowed to use the CFT-20 intelligence test, should they wish to do so, but only as supplementary information to the subject-matter performance of pupils when deciding (together with parents) which tracks students should enter. This is done at the end of the 4th school year and 6th school year. The test must be administered by "Testlehrer" and there is no systematic assessment of predictive validity.

In Lower Saxony, only the school psychologist or counsellor may select and use the available intelligence tests to test individually selected pupils for guidance and counselling purposes.

In Northrhine-Westfalia, tests may be used for career

Table 3: Number of published formal and informal tests, available in Germany (for Grade 7 and above).

Test purpose	Number of normed tests	Number of non-normed tests and test like material
Achievement tests		
General school achievement	1	-
General German language	2	-
Spelling	5	2 sets
Reading comprehension	1	2
Vocabulary	4	-
Grammar	-	2 sets
Mathematics / arithmetic	3	4 sets
Foreign language	6	8 sets
Science	2	12 sets
Social studies / history	-	2 sets
Combined achievement / Aptitude tests	2	1
Intelligence / apitude Tests		
Individual intelligence tests	4	-
Group intelligence tests - verbal	1	-
Group intelligence tests - non-verbal	6	-
Group intelligence tests - mixed	14	-
Special aptitude tests	11	-
Concentration / attentiveness Tests	6	-
Social attitude tests	3	-
Psychological "questionnaires" (e.g. anxiety, motivation, interests)	21	-

Source: K. Ingenkamp. Verzeichnis der deutschsprachigen Schultests. Stand Sommer 1984. In R.S. Jäger et al (Eds) Tests und Trends 4. Jahrbuch der Pädagogischen Diagnostik. Weinheim/Basel (Beltz) 1985.

guidance purposes in Hauptschulen and Realschulen as well as in comprehensive schools, but not in Gymnasia or vocational schools!

In Schleswig-Holstein, intelligence tests can be used "diagnostically" for career guidance, dyslexic problems, and behavioral problems.

In all cases, it is only qualified personnel who are allowed to administer the tests and interpret the results.

c) Dyslexic tests. In Hamburg special school personnel for dyslexia use existing tests (WRT6+ and RST8+) and develop new tests in order to decide on funneling students into special treatment programs. The LRS (reading writing weaknesses) teachers play a special role in the diagnostic testing of dyslexic children and in teaching them.

In Northrhine-Westfalia, new legislation abolished special grading practices for dyslexic pupils, identified on the basis of testing, except in comprehensive schools.

In Schleswig-Holstein, class teachers use WRT5+ (often together with CFT-20 intelligence test in accordance with the "standard" view of LRS as a special form of under-achievement.

d) Achievement tests. In Hamburg, the school authorities assume that some teachers use the publishing company tests but the authorities have no actual data. Some vocational schools "construct" tests in cooperation with the trade guilds in the Chamber of Commerce. The test must be recognized by the guild if it is to serve as a recognized exam.

Lower Saxony is similar to Hamburg.

In Northrhine-Westfalia, Gymnasium teachers are not allowed to use commercial achievement tests because of perceived problems of content validity. No funds for the acquisition of such tests are made available to any school. The vocational schools' use of tests is similar to Hamburg.

In Schleswig-Holstein, achievement tests are sometimes used when decisions about designation of pupils to special education have to be made.

In general, it can be seen that standardized tests are rarely used in German schools. There is no established culture of testing in schools and only limited empirical research. There is no consistent monitoring of ——————————→

achievement so that no one knows (but, presumably, some care) if achievement standards are rising, falling, or remaining constant.

Those supplying information to Professor Lehmann were:

Hamburg:
Dipl. Psych. A. Janowski, formerly Amt für Schule, responsible for research and testing (position now vacant), now University of Hamburg, Dept. of Psychology;

Dipl.Psych. Dr. P. May
Dipl.Psych. C. von Truchseß, both Dienststelle Schülerhilfe (a service institution of the Ministry of Education, involved in psychologically-based guidance and counselling and also in test development);

Lower Saxony:
Dipl. Psych. H. Diepenbrock, Schulpsychologischer Dienst, Hannover (community-based institution involved in guidance and counselling)

Northrhine-Westfalia:
Department heads of the Ministry of Education, Düsseldorf:
Herr Niel, responsible for Hauptschule, Realschule, Gesamtschule;
Frau Sebbel, responsible for Gymnasium;
Prof. Dr. Püttmann, responsible for vocational schools;

Schleswig-Holstein:
Dipl. Psych. Frau Greuer, Schulpsychologische Beratungsstelle Lübeck (guidance and counselling institution with 2 full-time staff primarily concerned with diagnosis and 2.5 temporary full-time ("ABM") staff for therapy.

C. THE NETHERLANDS +

1. The System.

The Netherlands created a national curriculum development center (SLO) only in 1975. This center produces models and syllabi. The regional Boards decide on which models / syllabi they will adopt. Universities also produce school curricula materials. Educational publishing houses develop curricula with different interested groups. Each school chooses the specific curricular materials it wishes to use within the framework of the "model" selected by the Board of Education. National examinations are a combination of internal assessment and the national written examination. In 1968 the Institute for the Development of Achieve-

---

+ Information for the Netherlands was supplied by Dr. Hans Pelgrum of the Department of Education of the University of Twente.

ment Tests (CITO) was founded. Its main aim was and is the development of mechanisms for the objective judgment of pupils' work.

Figure 2 presents the Dutch school system in diagrammatic form.

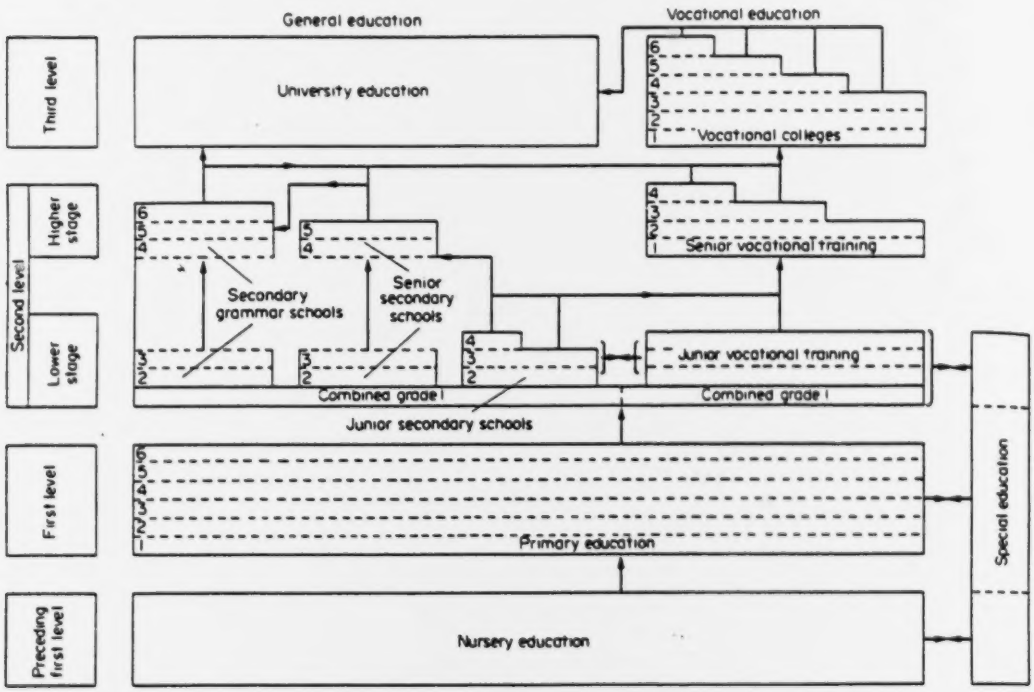


Figure 1
Structure of full-time education[a]

Secondary general education comprises four main types of school: Pre-university education (secondary grammar schools - VWO), junior (MAVO) and senior (HAVO) secondary schools, junior (including LTO, LHNO, and LAO types) and senior vocational training and vocational colleges, and miscellaneous types of secondary education - such as social training courses for young workers.

Add to this that the entire educational system can be divided into public, Catholic, Protestant, and secular schooling and one can begin to understand the complexity of the situation.

2. <u>What standardised tests exist?</u>

The publishing companies offer exercises in different subject areas (in tune with the text books they publish) but these cannot be considered to be standardised tests.

Since 1978 CITO has published criterion referenced tests in Biology, Physics, Chemistry, Mathematics, Dutch, English and French. These tests are meant to be used for formative education and diagnostic use by teachers during the teaching-learning process. They are based on an analysis of commonly used learning materials.

3. Who uses the tests for which purposes?

Approximately 50 percent of all Dutch schools ordered one or more of the CITO tests. Kremers (1982), however, found that only 15 percent of the teachers who ordered these tests also really used them (see Table 4). Besides that, most teachers used the tests for summative evaluation instead of formative evaluation (for which they were designed) and they modified the content of the test by rearranging items or combining subtests. Thio (1983) stated: "It proved that, in spite of efforts of promotion and information, the tests do not sell as well as was hoped by the CITO".

Table 4: The number of regular users of criterion referenced tests: Totals, per subject, per schooltype (percentages and frequency)
Source: Kremers (1982)

School type	Subject: Biology		English		French		Dutch		Mathematics		Total	
	%	F	%	F	%	F	%	F	%	F	%	F
Lower Vocational	13	12	9	12	0	0	9	10	23	29	13	63
General	28	52	5	7	13	19	10	12	23	38	17	128
Pre-university	38	3	0	0	13	1	0	0	0	0	12	4
Voc. + Gerneral	53	10	0	0	6	1	10	2	21	5	18	18
Gener. + pre-univ.	19	32	4	3	15	11	10	3	12	10	14	59
Total	25	109	6	22	12	32	8	27	20	82	15	272

Table 5 presents the reasons given for not ordering the CITO tests.

Table 5: Reasons for not using the ordered criterion referenced tests (figures are percentages of all teachers in each subject).
Source: Kremers (1982)

Reason	Subject					
	Biology	English	French	Dutch	Math.	Total
1. Course independence	71	69	40	69	65	64
2. Test cannot be administered in total	62	53	30	54	51	52
3. Tested goals only part of subject matter	62	49	42	60	39	51
4. Test only aimed at non-complex goals	39	38	57	48	35	42
5. Own course not in reference scheme	26	35	21	33	28	28
6. Tests too easy for our students	17	13	30	22	16	19
7. Tests too difficult for our students	14	11	10	20	14	14
8. Diagnosis can be done simpler	38	66	70	68	58	58
9. Administration costs too much time	38	48	61	37	57	48
10. Tests not suited for marking	19	31	37	16	30	27
11. Too many m-c items	35	35	52	46	49	43
12. Tests too short	19	27	24	35	6	21
13. Too much open answered	2	9	10	6	6	6
14. No time	59	63	70	61	68	64

Table 5 shows clearly that, on average, the most prominent reason for not using the tests was that teachers did not perceive them as suited to their own situation. Apparently, teachers want tests which match as closely as possible the content and teaching methods which they are using. This may be the explanation for the fact that in Table 5 the course-independency is of relatively low importance for the teachers of French; for this subject course-dependent tests were developed. There is some other indirect evidence for the use of tests from the Dutch involvement

in the mathematics and science studies of the International Association for the Evaluation of Educational Achievement (IEA). In these projects teachers were, amongst others, asked about their use of selfmade and other tests. Tables 6 and 7 present the results on the relevant questions.

Table 6: Percentage of teachers indicating that they use published tests.
Source: Pelgrum, Eggen, Plomp (1983
Second International Mathematics Study

	School Type			
Degree of use	Pre-univ. Senior Sec. N = 60	Junior Secondary N = 70	Junior Technical N = 57	Junior Domestic Science N = 49
Seldom / never	87	43	53	59
Sometimes	12	40	35	31
Often	-	14	12	10
No answer	1	3	-	-

Table 7: Percentage of teachers indicating that they use standardized tests.
Source: Pelgrum, Plomp (1986)
Second International Science Study

	School Type									
Degree of use	Pre-Univ. Sen.Sec.		Junior Secondary			Jun. Tech.	Jun. Dom.	Junior Agricult.		
	Phys.	Chem.	Biol.	Phys.	Chem.	Phys.	Biol.	Biol.	Phys.	Chem.
N=	45	48	42	35	43	52	41	24	19	23
Seldom/never	76	90	52	52	47	54	52	84	63	79
Sometimes	18	4	38	31	26	23	16	8	32	17
Often	4	2	10	14	23	21	25	4	0	0
No answer	2	4	0	3	3	2	7	4	5	4

Tables 6 and 7 indicate that the number of teachers using tests regularly is low. There are, however, considerable differences between school types. Currently we do not know what the reasons for these differences are. Janssens (1986) would seem to be right when he pointed out that there is very little research in the Netherlands into the use of achievement tests, and that more research is needed.

A Comment. In the Netherlands, achievement tests are used by teachers only. The use of other forms of standardised tests such as intelligence tests is similar to Germany where such tests are used by the school psychology and counselling units.

There is a move at present to institute a national assessment program of schools and students. Schools have shown interest in self-evaluation by comparing their results over time and with similar schools through the use of national assessment data. From informal observation during the IEA math and science projects Hans Pelgrum has suggested that school inspectors and school principals would be very interested in data on the achievement of certain schools and classes.

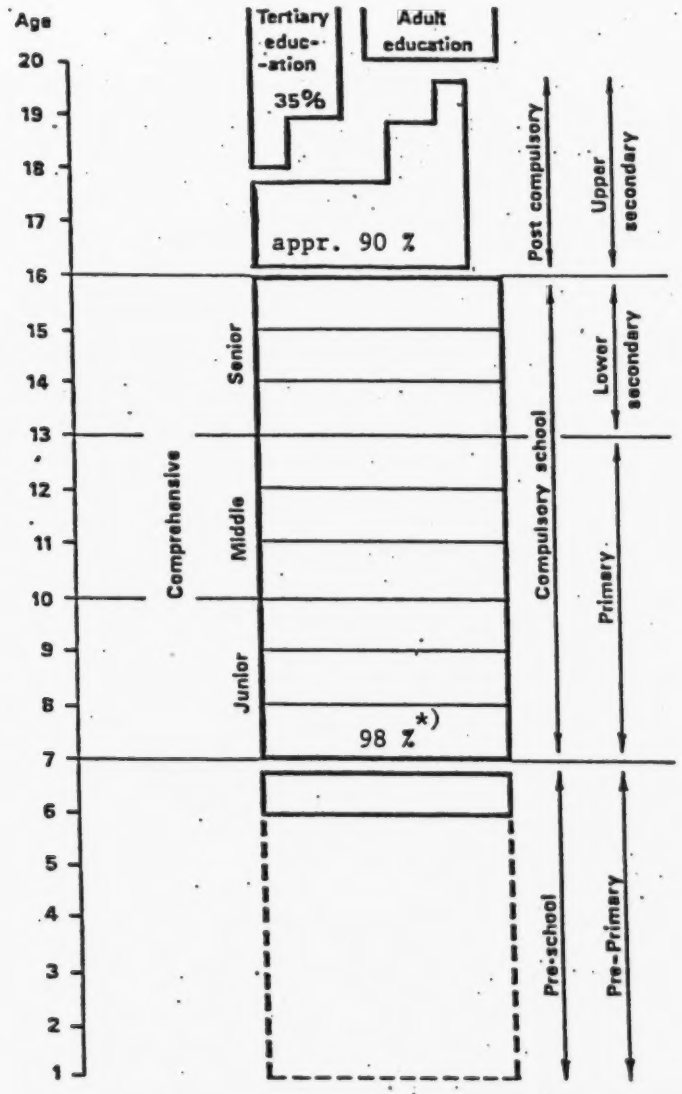
D. SWEDEN +

1. The System.

Secondary school is a term that ill fits the Swedish system of education. It has not been used in Sweden for many years. The Swedes speak of pre-compulsory (or pre-school), compulsory, and post-compulsory education. Figure 3 presents the structure of the Swedish school system.

+ The information in the section on Sweden was supplied by Prof. Sixten Marklund of the Institute for International Education at the University of Stockholm.

Figure 3. School Structure in Sweden in 1984.



*) Of every year cohort 1 % go to special schools for physically or/and mentally handicapped, and 1 % to private schools (compulsory level). The other 98 % join the regular basic school. Appr. 90 % of annual cohort continue to post-compulsory school and appr. 35 % of annual cohort to universities and other kinds of post-secondary schools.

Secondary school, for the purposes of this paper, is probably best defined as grades 7 to 12 (ages 13 to 18) or what is depicted in Figure 3 as senior comprehensive (or lower secondary) and the first three years of post-compulsory (or senior secondary). The school system is unitary in the sense that the same general and specific aims are pursued in the same kind of educational institutions all over the country. Thus, all those studying any given subject at the same level usually follow the same curriculum and have the same number of weekly periods. Courses and timetables are contained in a handbook (Läroplan) stating the overall aims of education as well as the aims and objectives of all subjects being taught, outlining the syllabus and giving the guidelines for each subject and discussing teaching methods and materials. However, in the final analysis, it is the teacher who undertakes the teaching in each classroom and so there is a certain amount of variation between classrooms on exactly what is taught and how it is taught. (The teachers "interpret" the Läroplan and there is, of course, adaptation of the classroom work to the students' individual interests and aptitudes).

However, marks are given for each student's work in different subject areas. These marks are awarded by the teacher in a specific subject area. But to guarantee, as far as possible, that the marks have the same value all over the country (marks are on a 1 - 5 scale with 5 being high) standardized tests are used.

To quote from Marklund (1985): "The marks given to all students in the same grade studying the same subject and, where alternatives exist, taking the same course - "general" or "special"- should be spread out by the mark-giving teachers according to an approximate normal distribution, as shown below. It is important that this normal distribution of makrs refers to the whole country. Single schools and classes usually spread differently.

Mark	1	2	3	4	5
%	7	24	38	24	7

In the compulsory school (grades 1-9) the actual distribution follows these figures fairly well for the nation as a whole. In the post-compulsory school (upper secondary) the actual distribution of marks, due to the students' choice of specialization, has gone a little "upwards", with the national means around 3.4

or 3.5. (How marks are given is further described in Appendix II "The Assessment Process").

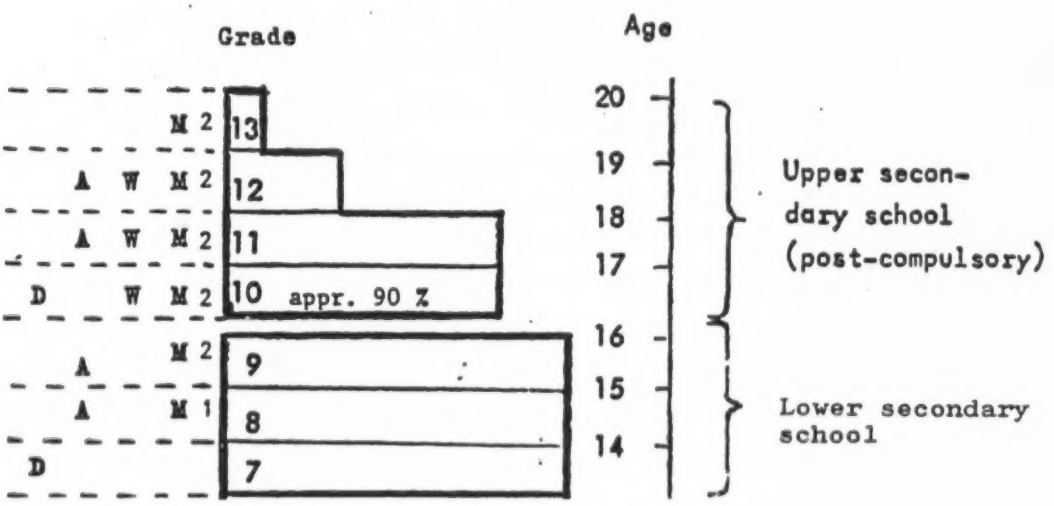
The mark 3 denotes the mean accomplishment of the total population of students in the whole country doing the same course, as explained above. Thus the mark received by any individual student expresses to what extent he or she has succeeded in relation to that population, in achieving the aims and objectives set for the subject in question. Obviously, no marking system can be perfect, in the sense that it always does absolute justice to each individual student. However, by means of the regular nation-wide application of standardized achievement tests based on objective techniques, it has proved possible to go a long way towards stabilizing the marking system and eliminating variations due to change.

In the primary stage students do not get any marks. At this level local school authorities decide on other forms of information to parents and students, usually by oral reports but also by written non-formal reports. Marks are then given at the end of grade 8, and thereafter at the end of each term, i.e. twice a year, throughout grade 9 of the Basic School and the whole of the Upper Secondary School. Marks given at the end of the autumn term indicate the level of achievement reached during that term, whereas spring term marks are based on the student's performance during the whole of the academic year."

Figure 4 presents the testing and assessment procedures at the different grade levels in "secondary" school.

Figure 4. Testing and assessment in different grades

D = Diagnostic tests, voluntary
A = Standardized achievement tests, compulsory in grades 10-12, voluntary in grades 3, 6, 8 and 9 although used by 90 % of the teachers in these grades
W = Written tests, compulsory
M1= Marks given at the end of the school year
M2= Marks given at the end of the autumn term and at the end of the school year



2. What tests exist?

There are many standardized tests that are used in school. They can be sub-divided into achievement tests and diagnostic tests. The first assess the achievement, group and individual, of the total population in any one subject at any one grade level; their purpose is to enable the teacher to compare the performance of his or her own class with that of the total population and to adjust his or her marking scale according to the outcome of the testing.

The second kind, individual diagnostic tests, are given at the beginning of a learning unit or set of units in order to provide a detailed profile of the students' skills and knowledge. The outcome is meant to help teachers and students to draw up a study program which will meet the specific needs of individuals and groups, or of the class as a whole.

Table 8 presents a summary of the achievement and diagnostic tests used in schools in 1983.

Table 8: Diagnostic and Achievement Test used in grades 7 to 12 in 1983.

Subject	Grades							
	Lower Sec.			Upper Sec. 3-4 year streams			Upper Sec. 2 year streams	
	7	8	9	10	11	12	10	11
Swedish	D		A			A		A
Math	D		A	D		A	D	
English		A			A			A
French/German			A		A			
Chemistry				D	A			
Physics				D	A	A		
Mechanics							D	

Until 1982 it was a section of the National Board of Education that was responsible for the construction of these tests but since then the task has been decentralized to educational research institutes (Malmö for Swedish, Gothenburg for foreign languages, Stockholm for mathematics, Umea for Science, etc.).

Appendix III presents a summary (Marklund, 1985) of the requirements and construction of the tests, including the way in which a "quick standardization" is carried out.

The reliability of the standardized tests tends to be over 0.90 (KR 20) but some are in the 0.80s. The validity of the tests is usually estimated by means of simple correlations with the teacher's marks (range .5 to .9). The face validity is checked by having experienced teachers estimate the relevance of the tests in relation to the Läroplan. In Swedish written composition, the teachers are given examples of essays which have been judged to be "poor", "average", and "good".

3. Who uses the tests for which purposes?

All teachers, as has been seen above, use standardized tests partly voluntarily (diagnostic tests and achievement tests in lower secondary grades) and partly because it is required by the National Board of Education. The school principals also use the results

in the sense of helping plan the future courses (in terms of the relative strengths and weaknesses of the school in achievement within various subject areas).

It is only recently that the Swedish Parliament has accepted the idea of a national program of evaluation (Nationallt program för utvärdering). This is a form of national assessment where the content of the Läroplan will be closely adhered to. The results will be used by national, regional, and local authorities. However, single student and class results will not be published.

CONCLUSION

It would seem as though teachers do not use standardized tests very much of their own accord. Only in one of the four countries, Sweden, are tests used a lot but this is because they are imposed by the government for the purpose of calibrating teachers' marks in a system of continuous assessment because there are no formal examinations.

The main reason for not using tests is that the content is too general or, put another way, the tests are not tailor-made for what the teachers have been teaching.

To paraphrase Tyler (1986), teachers (and parents) need to know which children have learned what they have been taught and what has each child not learned that he should have learned so that corrective action might be taken; in other words, criterion-referenced tests for formative (and, occasionally, summative) evaluation purposes. Whereas the teacher needs information on each child, the school principal needs to know about the progress of learning in each classroom so that assistance can be given when needed. In a decentralized system of education, this can be very helpful for the purpose of setting school goals (in staff-conferences). District officers do not need such detailed information as teachers, parents, and school principals. The district personnel need to know about the different proportions of children having difficulties (and, of course, proportions succeeding on different forms of objectives). Breakdowns of achievement by school type, sex, urban - rural or on other variables thought to be important, is what the district officials need. The state, regional and national authorities, are responsible for policy.

They need to know what children in their area are learning, what learning is expected of them at various stages of their development and what progress the children are making and what problems they are encountering.

England has its A.P.U. but exactly how useful it is for any of the above purposes is not clear. Germany would appear to have basically nothing in that the only standardized tests are those used by school psychologists and the guidance and counselling personnel; the teachers use quizzes and the state and national authorities have no systematic empirical evidence by which to judge standards of achievement either for each state or for the nation as a whole.

The Netherlands has its C.I.T.O., but even so only 15 percent of all teachers actually use tests. But, national assessment will soon begin.

Sweden has its standardized tests, and it will soon begin its national assessment.

Some years ago it was thought that item banks would be the answer. With carefully constructed item banks (with Rasch scale values attached to each item) it would, at least in theory, be possible for any teacher to sit at a terminal and screen and review and select items to test exactly what she had taught last week. Because the scale values were known, it would be possible for the teacher, after testing her students, to have not only information on how well or poorly each student in the class was performing on each item but also how the class as a whole compared with other similar classes in the region or nation. Probably the most advanced system is the Ontario one (for which the Ontario Institute for Studies in Education was producing the prototype); but, to my understanding, this is not operational. In an O.E.C.D. meeting that I attended recently there seemed to be doubt about item banks operating through terminals actually working. Rather, it was said, teachers prefer to have books of items that they can choose from. Videodisc, it is suggested, may replace the books of item.

The formative tests (and remedial materials) produced by the Korean Educational Development Institute (K.E.D.I.) for Grades 7, 8, and 9 as supports for the Mastery Learning system

would appear to have worked very well in increasing achievement nationally. These were tailor-made to test the pre-specified content of the learning units.

It would appear that more research is needed in those countries where standardized tests are used into exactly how the teachers use the test results once they have them.

However, at the state level some U.S. states may wish to look more closely at the practices of England, the Netherlands or Sweden.

REFERENCES

Booth, C., (1985) United Kingdom: System of Education. In: Husén, Torsten and Postlethwaite, T.Neville, International Encyclopedia of Education. Vols. 1-10. Oxford, Pergamon (pp. 5351-5359).

Gipps,C. et.al.,(1983) Testing Children. London, Heinemann Educational Books.

Janssens, F.J., (1986) Toetsgebruik in de onderwijspraktijk (Test use in education), Tijdschrift voor onderwijsresearch (Journal for Educational Research), 11, 1.

Kremers, E.J.J., (1982) Gebruikers- en behoeftenonderzoek: project leerdoelgerichte toetsen.(Survey of users of tests and their needs: a project), Arnhem, Cito, documentatiereeks Nr. 68.

Marklund, S., (1985) Education in Sweden: Assessment of Student Achievements and Selection for Higher Education. Stockholm, mimeo.

Nijhof, W.J., and Streumer, J.N., (1985) Netherlands: System of Education. In: Husen, Torsten and Postlethwaite, T. Neville, International Encyclopedia of Education. Vols. 1-10. Oxford, Pergamon (pp. 3501-3510).

Pelgrum, W.J., Eggen, Th.J.H.M., Plomp, Tj., (1983) Tweede Wiskunde Project: Beschrijving van uitkomsten (Second Mathematics Project: Report of the outcomes). T.H.-Twente, Toegepaste Onderwijskunde.

Pelgrum, W.J., Plomp, Tj., (1986) Second International Science Study: Beschrijving en analyse van uitkomsten (Report of the analyses of outcomes). T.H.-Twente, Toegepaste Onderwijskunde.

Thio, K.D., (1983) Vijftien jaar Cito, vijftien jaar 'Wiel Solberg' (Fifteen years Cito, fifteen years 'Wiel Solberg'. In Weeda e.a., Examens in discussie; een bundel opstellen voor J.W. Solberg. Groningen: Wolters-Noordhof.

Tyler, R.W., (1986) Changing Concepts of Educational Evaluation. International Review of Education, 10, 1. Oxford, Pergamon.

Very Recent Developments in Tests in England

Ray Summer

These stem from the change in modern language teaching, which traditionally was based on learning grammar and vocabulary and focussed largely on reading with comprehension and writing correct prose; pupils' speaking was assessed in the external examinations though a 10 minute 'oral' following a brief preparatory read through a passage or scrutiny of a picture. Large proportions of pupils abandoned languages (mainly French and German, some Spanish and Italian) after 2 or 3 years of compulsory study in favour of other options at the age of 14+. Language teachers then took up 'graded objectives' aimed at developing communicative competence (and retaining more pupils in later years).

Perhaps as many as a half of the 104 authorities in England and Wales have Graded Tests of Modern Languages; (various titles, e.g. GOLF, i.e. Graded Objectives in Learning French). Early stages are oral/aural and so are the tests; later stages involve functional literacy. General practice is to devise fresh tests annually, though now that schemes have run for some years (up to 10), re-cycling is being practiced. Standardisation is via consensus and procedures such as guidelines and comparison with tape-recordings illustrative of pass/fail levels.

Clearly, these schemes are firmly related to curricula; hence, schools relinquish the feedom to vary from the 'core', which will be assessed, but in other respects, retain their autonomy regarding curricular content, skills and style of teaching. In principle, pupils can take a test at any level as and when they are thought to be sufficiently proficient. In practice, there is a marked tendency for the tests to be used as end-of-year assessments; 1st year given level 1, etc. Problems in dealing with the logistics of individual oral testing have been reported; additionally, the organisational difficulties of coping with teaching classes of pupils who rapidly differentiate into different levels have been said to inhibit 'testing when ready'. A further point is that teachers who have not been on the test panels are somewhat unsure of the requirements; in other words, the test provides a better definition of course objective than the formal statements of aims.

As implied previously, test construction is not technically sophisticated. Teachers on the local panels choose or devise stimulus materials and simulated situations and may try these out on their own pupils. It is unlikely that trial data on items is analysed or that standard-setting procedures, other than broad consenus, are used; no test reliability figures will be calculated or sought. However, there will be training in judging performance; i.e. listening to tapes and there may be cross-moderation, i.e. visits to schools doing the tests by a senior assessor.

Other Graded Tests. The modern language model has been followed, to some extent, by (i) other localities for certain subjects, and (ii) the external examining bodies. Hence, there are schemes for Mathematics individualised learning which incorporate topic tests, and both topics and tests are defined by level. Similarly, Science tests have been devised and marketed, for practical science processes (often called skills, e.g. reading a measuring device).

A notable point is that several schemes currently being developed by external examining bodies do not utilise tests at all. These schemes are called 'graded assessments' and so involve teachers as judges very heavily as compared with standardised tests, where whatever is judged is in the remit of the test constructor. The dividing line between a test and a product from a pupil submitted for judgement is not, however, all that pronounced. In one scheme, 'eight workpieces have to be approved at a certain level to qualify for the Level award, and these might be done quickly by some pupils whilst others could take up to 8 weeks'. This example illustrates that whilst tasks may be standardised conditions may vary greatly. The Mathematics scheme is heavily dependent on the curriculum materials devised for the Levels and on training the teachers to use assessment criteria, some of which concern processes.

Similar schemes are under way for Science, Craft, Design and Technology (CDT), and English. In English, the definition of level has given way to the idea of breadth; in other words, a menu of competences is available for assessment and, furthermore, teachers may work as they choose in preparing pupils for an assessment; there is an implication in this subject, that more units passed corresponds to a higher level of competence. This is likely to be formalised if the examining body agrees to a trade-in procedure

whereby its own graded awards are granted when pupils' graded work has been verified (by inspection). Linking with an external body which issues nationally accredited awards is a considerable incentive to the schools.

The Assessment Process

Sixten Marklund

Classroom Observations

The teacher's main task is, of course, to aid the students in their personal development and to help them acquire the skills and knowledge defined in the aims laid down in the Curriculum.

This entails continually assessing the students' work and keeping them informed of their progress. Teachers are therefore advised to observe each individual's performance within the class and to record their observations from time to time.

All performances must be taken into account, and the teacher must be on his guard against paying too much attention to results that are easier to assess than others. It is particularly important to take proper account of oral proficiency, in the mother tongue as well as in foreign languages, since this most important ability cannot at present be easily measured by means of objective techniques.

The Upper Secondary School class used to be visited occasionally by a subject expert. These experts study the work in progress and discuss it with heads, teachers and students, both in conference and privately. They are thus able to form a good overall picture of all school activities concerning their subjects and of the general standard of skills and knowledge achieved in different schools, as well as to give advice on teaching methods and evaluation. In the Basic School the same functions are performed by other categories of inspectors and advisers.

Written Tests

The teacher keeps a record of each student's performance in all written tests taken during the evaluation period. In the Upper Secondary School all compulsory test papers are filed so as to be available for principals and visiting inspectors. By examining the papers, they are able to see if the marking principles applied by the teacher tend to be more lenient or severe than the average and is thus in a position to assist teachers in their endeavour to attain a high degree of uniformity in assessing the students' work.

Final Assessment

Towards the end of the term, the teacher surveys all the evaluation data collected as described above, and ranks the students from top to bottom according to their individual level of ability, giving each a mark on the five-point scale. These marks are preliminary and may have to be adjusted. As stated above, the Curriculum emphasizes that the students' standard of performance within the class must be given proper weight in relation to their results on written work. In the job of assessing the students' overall standard, the teachers will find their task greatly facilitated if they have kept a running record of their classroom observations.

The main function of the standardized test is to be instrumental in achieving the highest possible degree of uniformity in the marking system. A detailed description of the procedure to be followed is contained in the Curriculum. A brief summary is given below.

First the teacher calculates the mean of the prelimenary marks and records their distribution over the five-point scale. Then he compares these data with the mean and distribution of marks obtained by the class in taking the nationally standardized test. If the two means are identical, or if the difference between them does not exceed $\pm$ 0.2 (which used to be seen as an acceptable tolerance for chance influences), the teacher can conclude that the preliminary marks indicate the standard of the class correctly in relation to that of the total population. If the two distributions also coincide more or less completely, the preliminary marks can be taken as final.

Each teacher delivers the marking documents to the headmaster's/ headmistress's office, all the relevant data are arranged and recorded in such a way as to facilitate comparisons between classes and within each class. This material is available at a meeting, called a class conference, which is attended by the head, and all the teachers taking the class in question for one or more subjects. The purpose of the class conference is to take final decisions on the means and distributions of marks. Comparisons are made between the achievements of different classes in the same subject. A teacher who wants to retain noticeable differences between test results and preliminary marks has to convince the class conference that there is a valid reason for doing so.

The adjusted means and distributions of marks for those subjects in which standardized tests are taken, are used as guidelines for adjusting the means and distributions for other subjects. This principle is based on the well-known fact that within a class the means and distributions have as a rule a fairly high degree of correlation, regardless of subject.

The dividing up of the marking procedure into two steps, one for preliminary marks and one later for final marks, is important. The class conference between these two steps aims at making single marks for single students comparable all over the country. This way it has become possible to base the selection for higher studies on secondary school marks instead of university entrance examinations.

Standardized Tests: Requirements and Construction

Sixten Marklund

Requirements

All standardized tests have to fulfil certain requirements. They have to be valid in the sense that they actually measure the skills and knowledge defined in the aims as accurately as possible.

In principle, the tests should cover all essential aims as laid down in the Curriculum. This is not possible, however, because so far no sufficiently economical and efficient techniques exist for the testing of some aims, e.g. oral proficiency.

Diagnostic tests should assess as many relevant learning objectives as possible, otherwise they fail to indicate what special measures should be taken to adjust the learning process adequately. Achievement tests can be less detailed because, in the case of nation-wide reference group, there are usually high correlations between data obtained by measuring different abilities within the same subject. On the other hand, if an important ability is never subjected to testing there is risk that it may be neglected also in the training programme.

Achievement tests have to differentiate clearly betwwen testees, ranking them according to their performance from top to bottom, with a high degree of reliability. The all important thing is to ensure that as far as possible the marking of these tests is uniform throughout the country, leaving no room for personal preference or bias on the part of the marker. This end is achieved either by using entirely objective techniques based on the multiple choice principle or, where this is impossible or considered undesirable, by reducing the influence of subjective judgement to such an extent as to make it negligible.

Construction

A section of the National Board of Education has until recently been responsible for the construction and distribution of all standardized tests in regular use, and for instructions as to their

application. Now the test construction is taken over by educational research institutes at the universities. For each subject there is a steering committee consisting of subject experts as well as experts on psychology and psychometry. In order to ensure the necessary feedback from schools to the test makers, some committee members are active teachers. The committee is responsible for the analysis of aims and objectives necessary to secure test validity for the national school system, and for the testing policy to be adopted by the schools, i.e. establishing principles for the choice of elements or content areas to be tested and for the structure of the tests.

The test constructing institutes commission some subject experts, who are as a rule active teachers, to construct test items along the adopted lines. The result of their work is submitted to the committee, who makes such revisions as are deemed appropriate. The revised version is then tried out in a number of schools. The text experts used to be about 150 altogether, most of them acting for short periods and temporary meetings.

The testees' answers are recorded and a detailed item analysis is made by the steering committee on the basis of data obtained by computerizing the test results. Items that have proved to be unsatisfactory as to reliability are scrapped or altered. Where computerizing is not feasible, other measures are taken to attain the highest possible degree of reliability.

In due course, the finalized version of the test battery is sent to all schools concerned, together with detailed instructions on testing procedures. The tests for the Upper Secondary School are compulsory but not those for the Basic School, where, however, about 90 percent of the teachers use them. The latter tests are used repeatedly over a period of some years so they have to be kept confidential, whereas new tests are, at present, constructed annually for the Upper Secondary School. After they have been used, they may be published and discussed openly.

In recent years a simplified method of standardization has been practiced. This method, called "quick standardization", means that the tests are not at first tried out on a representative sample of testees before they are used. The first version of the

test, composed and darefully discussed by experts and steering groups, is applied directly. Replies from a representative sample of testees are then immediately collected. Norms on a five point scale of the results are then developed by the test constructors and quickly distributed to all schools, where the teachers - after having waited for these norms during a couple of weeks - now can record the test results of their students.

The advantages of this "quick standardization" are obvious. The try out round can be abolished, which saves time and money. The risk of getting poor item in the instrument has proved to be minimal. A prerequisite certainly is, that the test construction experts and the steering committees are experienced test makers with a good knowledge of how different kinds of test items and instruments work on different levels of school and different levels of student ability.